

PREPRINT - Seepanomwan K., Caligiore D., Baldassarre G., Cangelosi A. (2013). Modeling mental rotation in cognitive robots. Adaptive Behavior, 21(4), 299 - 312

Modeling Mental Rotation in Cognitive Robots

Kristsana Seepanomwan¹, Daniele Caligiore²,
Gianluca Baldassarre² and Angelo Cangelosi¹

Abstract

Mental rotation concerns the cognitive processes that allow an agent to mentally rotate the image of an object in order to solve a given task, for example to say if two objects with different orientations are the same or different. Here we present a system-level bio-constrained model, based on neurorobotics, that provides an embodied accounts of mental rotation processes relying on neural mechanisms involving motor affordance encoding, motor simulation, and the anticipation of the sensory consequences of actions (both visual and proprioceptive). This model and methodology are in agreement with the most recent theoretical and empirical research on mental rotation. The model was validated through experiments with a simulated humanoid robot (iCub) engaged in solving a classical mental rotation test. The results of the test show that the robot is able solve the task and, in agreement with data from psychology experiments, it exhibits response times linearly dependent on the angular disparity between the objects. This model represents a novel detailed operational account of the brain sensorimotor mechanisms that might underlie mental rotation.

Keywords

Mental rotation, computational robotic model, neurorobotics, neural mechanisms, affordances and forward models, parietal/premotor/prefrontal cortex.

1 Introduction

Since it was first described by Shepard and Metzler (1971), mental rotation has attracted enormous research interest in the field of cognitive psychology. This is in part due to the attempts to understand why object comparison using imagery seems to obey the same physical principles as overt rotation, considering that humans are capable of using imagery that is not limited by the laws of physics (Kosslyn, 1994). In a typical mental rotation task, human participants are asked to make a decision on whether two objects presented with different rotational orientations are identical or a mirror version of each other. The results show that the response times (RTs), as well as the errors, of the participants' answers, are highly dependent on the angular disparity between the two stimuli (Shepard & Metzler, 1971; Wexler, Kosslyn, & Berthoz, 1998). In particular, participants show RTs that linearly increase with the disparity angle between the orientations of the objects. The number of errors also increases with the disparity increment. The most accredited explanation of these results is that the participants might rotate a “mental” image of one object until its orientation matches the one of the other object (Kosslyn, 1994). Once mentally rotated,

the participants can ascertain if the two objects are identical or not.

Early attempts to explain brain mechanisms underlying mental rotation processes relied upon a visuo-spatial perception hypothesis (Shepard & Metzler, 1971; Corballis & McLaren, 1982). According to this view, mental rotation is performed on the basis of processes mainly involving the internal manipulation of the visual and spatial features of objects. This view makes the prediction that these processes mainly implicate brain areas underlying visual and spatial perception. Contrary to this, recent behavioural and neuroscientific evidence indicate also an important involvement of motor processes, aside the perceptual ones. In this respect, several behavioral works show interferences between action planning/

¹Plymouth University, Plymouth, UK

²Laboratory of Computational Embodied Neuroscience, Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche (LOCEN-ISTC-CNR)

Corresponding author:

Kristsana Seepanomwan, Plymouth University, Drake Circus, Plymouth, PL4 8AA, UK

Email: kristiana.seepanomwan@plymouth.ac.uk

execution and mental rotation processes (Wexler, Kosslyn, & Berthoz, 1998; Wohlschläger & Wohlschläger, 1998; Wohlschläger, 2001). In a typical experiment, participants are asked to perform a classical mental rotation task (Shepard & Metzler, 1971) while performing a manual rotation on a custom joystick in both congruent and incongruent conditions with respect to the direction of rotation of the mental image. The results show that RTs (and error rates) are faster (lower) when the direction of the two rotations (manual and mental) is congruent, whereas they are slower (higher) when they are inconsistent (Wexler, Kosslyn, & Berthoz, 1998; Wohlschläger, 2001). This supports the idea that motor processes play a key role in mental rotation as otherwise it would be difficult to explain why the production of overt motor actions can interfere with mental rotation only when the two are incongruent.

Single cell recordings in the monkey's motor cortex also supply direct neural evidence for the involvement of motor processes in mental rotation (Georgopoulos et al. 1989). In humans, a number of neuroscientific studies using different research techniques, such as transcranial magnetic stimulation (TMS), event-related potentials (ERPs), and functional magnetic resonance imaging (fMRI), show an involvement of lateral and medial premotor areas (lateral premotor cortex/precentral gyrus and supplementary motor area) during mental rotation (Lamm et al. 2007; Richter et al. 2000). The fMRI study of Richter and colleagues (Richter et al. 2000), for example, shows a significant correlation between the hemodynamic response in lateral premotor areas with the response time of participants involved in the classical Shepard and Metzler mental rotation task (Shepard & Metzler, 1971). This result suggests that mental rotation is an imagined (covert) object rotation action rather than an image transformation relying exclusively upon visuo-spatial processing. This claim has been further confirmed by other studies (cf. (Wohlschläger, 2001; Lamm et al., 2007; Lamm, Fischmeister, & Bauer, 2005).

Importantly, despite these consistent results about the involvement of motor processes during mental rotation, we still lack a comprehensive hypothesis of the specific brain mechanisms involving motor simulation that might underlie mental rotation processes. One proposal that might help to explain the role of premotor areas during mental rotation pivots on the concept of affordance (Gibson, 1979) and its behavioural manifestations (Tucker & Ellis, 2001), brain correlates (Rizzolatti & Craighero, 2004), and models (Caligiore et al., 2010; Fagg & Arbib, 1998). According to this perspective, affordances are the possible actions that objects and the environment offer to a certain agent. In particular, the visual

presentation of objects triggers the activation, within the parietal-premotor circuits, of internal representations (the representations of affordances) needed for the on-line guidance of actions over them (Grafton et al., 1997; Grèzes & Decety, 2001). In this respect, the activation of affordance representations might be involved in the mental rotation processes as in brain it plays key role in the first stage of motor preparation.

Another hypothesis on how motor areas might participate in mental rotation comes from the theories (Grush, 2004), neuroscientific evidence (Miall, 2003), and computational architectures (Wolpert & Kawato, 1998) on motor control based on forward models. This perspective suggests that preparatory/planning covert motor processes play a key role in the mental simulation and understanding of the environment, and involve the same brain motor areas involved in overt action execution. This view would suggest that mental rotation involves the same motor areas and mechanisms used in the physical execution of active rotations of objects (e.g., manual rotations), and the imagined anticipation of their sensory consequences.

So, both views would give important indications on the possible involvement of motor areas in mental rotation phenomena. However, they would both still be limited in that mental rotation is a complex process requiring the coordinated operation of several distinct elemental cognitive processes. These processes include (Lamm et al., 2007): (a) stimulus encoding and mental image generation, (b) planning and execution of the mental rotation, (c) comparison (matching) of the rotated stimulus with the target stimulus, and finally (d) execution of the same/different response.

In this article we propose a system-level computational model suggesting a specific operational hypothesis on how the information processes taking place in brain sensorimotor areas might interplay to perform mental rotation. This hypothesis first draws ideas from the affordance and forward model view introduced above and integrates and specifies them to make them applicable to the explanation of mental rotation. Second, it introduces some additional elements to allow the implementation of not only the processes “a-b” indicated above (mental rotation proper), but also “c-d” (control and exploitation of mental rotation processes).

To this purpose, the model leverages on the computational model “TRoPICALS” (Caligiore et al., 2010; Caligiore et al., 2012) developed to study affordance compatibility effects (Tucker & Ellis, 2001). TRoPICALS is a good starting point to design a model on mental rotation as it reproduces some key functions of the parietal-premotor circuits, crucial for stimulus encoding and extraction of object affordances (process “a”) and also includes important

features of the prefrontal-premotor circuit pivotal for managing other key aspects of mental rotation (processes “c” and “d”). However, it cannot perform mental simulations as it lacks the needed feedback circuits. In this respect, to address the core mental rotation process (process “b”) the model proposed here enhances the functions of TRoPICALS by developing some key new features. First, it is endowed with premotor-parietal feedback loops that allow it to implement mental simulation and sensory prediction based on forward models. Second, it is endowed with enhanced parietal functions for encoding somatosensorial information important to elaborate anticipated proprioceptive signals. Third, it is endowed with an improved visual and motor system allowing it to scale up to more realistic 3D environments and robotic setups.

The rest of the paper is organized as follows. Sec. II discusses the main features of the model, the learning algorithms used to train it, and the robotic set up used to validate it. Sec. III presents and discusses the results. Finally, Sec. IV drives the conclusions.

2 Methods

2.1. The simulated mental rotation experiment: participant, stimuli and task

Figure 1 shows the simulated humanoid robot iCub (Tikhanoﬀ et al. 2008) we have used as a participant to model psychological experiments on the embodiment bases of mental rotation. It replicates the same body and control scheme of the real iCub robot (Sandini, Metta, & Vernon, 2007), which is an open source robotic platform built for studying cognitive development in humans. iCub looks like a human 3-5 years old child, in great part designed to be constrained on humans’ body structures and movements. Thanks to these features, the iCub platform is widely used as benchmark cognitive robotics tool in many robotics laboratories (Cangelosi & Schlesinger, in press).

The iCub simulator provides visual perception via simulated cameras and can perform actions corresponding to specific motor commands. Each arm of the iCub has 16 joints. Here we use the joint number 5 of the right arm affecting the robot wrist’s angle. If the robot holds an object with the right hand, rotating the wrist will change only orientation in the object plane. During the mental rotation task the model has to compare two visual stimuli having different orientation.

The stimuli are colored in red to make easier their detection by the iCub camera. The edge detection method is used as an early visual processing stage. The image is centred on a single object, and the red

color filter is applied. The edges of the object are extracted with the Canny edge detection technique (Canny, 1986), using the OpenCV library. The output from the edge detection process is converted as activation level to the neural network, input units at the beginning of the simulation. The left object is considered to be the target stimulus, whereas the right one represents the current stimulus which has to be mentally rotated (left stimulus to PP, right stimulus to PFC_1, see figure 2). The eyes’ position of the iCub is fixed, with the object centered on the fovea throughout the experiment. Regarding the motor response, the iCub’s wrist angle can rotate in the range of $[-90^\circ; 90^\circ]$. Counter-clockwise orientations are indicated by positive values, while clockwise orientations are indicated by negative values.

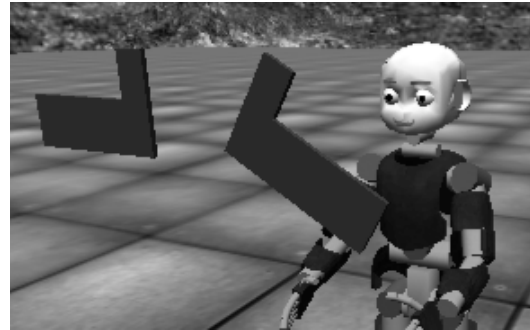


Figure 1. The iCub simulator, its environment and a sample pair of stimuli.

During the experiment pairs of target-current object images having different orientations are used. The objects are displayed in the space in front of iCub (Figure 1). For the training, the rotation of the comparison object is varied by 30° per pair, so that each stimulus could assume seven orientations ($-90^\circ, -60^\circ, -30^\circ, 0^\circ, 30^\circ, 60^\circ, 90^\circ$). During the process of affordance training, only one stimulus is shown in the left position, with the experimenter varying the orientation of the object and assigning a corresponding target position for the robot’s wrist angle. In the testing session, two stimuli are displayed in the left and in the right positions. In each trial, the rotation of the left image is systematically varied, while the right one is presented with several degrees orientations.

After training, the generalization ability of the model is tested using 196 pairs of stimuli, which are supplied in sequence to the model. The experiment has been repeated 10 times to test the consistency of the model. Each time the pair of stimuli is changed, the model internally rotates the left stimulus to match with the right one and produce an answer. Three types of information are recorded during the experiment: the response times (RTs) which is the result of a neural dynamical competition (see sec. 2.2.2 and cf. (Caligiore et al., 2010; Erlhagen & Schöner, 2002);

the answer for the current mental rotation task (see sec. 2.2.2); the successful degree of rotation (the maximum number of rotation cycles is set to 10, as in some cases the model cannot rotate the image to a preferred orientation at the first cycle, thus requiring extra rotations). When the number of rotation cycles is equal to 10, it indicates that the model cannot correctly perform the mental image rotation of the left stimulus and will be forced to do the next step (matching process) by using the last image.

2.2. Neural architecture, simulated mental rotation, learning phase

2.2.1 Neural Architecture

The neural network model (Figure 2) proposed in this article suggests an operational hypothesis about the interplay of the visual and motor neural processes during mental rotation. To accomplish this aim it extends some features of the TRoPICALS model (Caligiore et al., 2010). The architecture of the model is shown on Figure 2. It consists of four main parts corresponding to specific areas of the brain mainly involved during mental rotation tasks (Lamm et al., 2007; Richter et al., 2000): the parietal cortex (PC), the premotor cortex (PMC), the prefrontal cortex (PFC), and the primary motor cortex (M1).

These cortical areas are constituted by neural maps activated using population code methods (Pouget,

Dayan, & Zemel, 2003). The population code theory claims that information (e.g., on stimuli and actions) is encoded in the brain on the basis of the activation of populations of neurons organized in neural maps having a broad response field. In particular, each neuron responds maximally to a certain value of the variables to encode, and then progressively less intensely to values (based on a Gaussian function).

PC is formed by two distinct areas: the posterior-parietal cortex (PP) and the somato-sensory cortex (SS). The neurons of the PP map (32 x 32 neurons) encode the shape and the orientation of the object that has to be mentally rotated (Rizzolatti & Craighero, 2004). The neurons of the SS map (31 x 100 neurons) elaborate the proprioceptive signal related to the robot wrist orientation (Caligiore et al., 2010). The PMC region is formed by 2 neural maps PMC_1 (31 x 100 neurons) and PMC_2 (10 x 20 neurons), encoding motor programs related to different arm parts (Rizzolatti & Craighero, 2004). PMC_1 neurons encode the wrist posture of the robot corresponding to the object orientation encoded in PP. PMC_2 neurons encode the hand posture that the robot produces to accomplish the mental rotation results (i.e. to indicate if two objects are same or different). The PFC also consists of 2 maps, respectively implementing the working memory (PFC_1, 32 x 32 neurons) and the matching process area (PFC_2, 64 x 64 neurons) (Fuster, 2001).

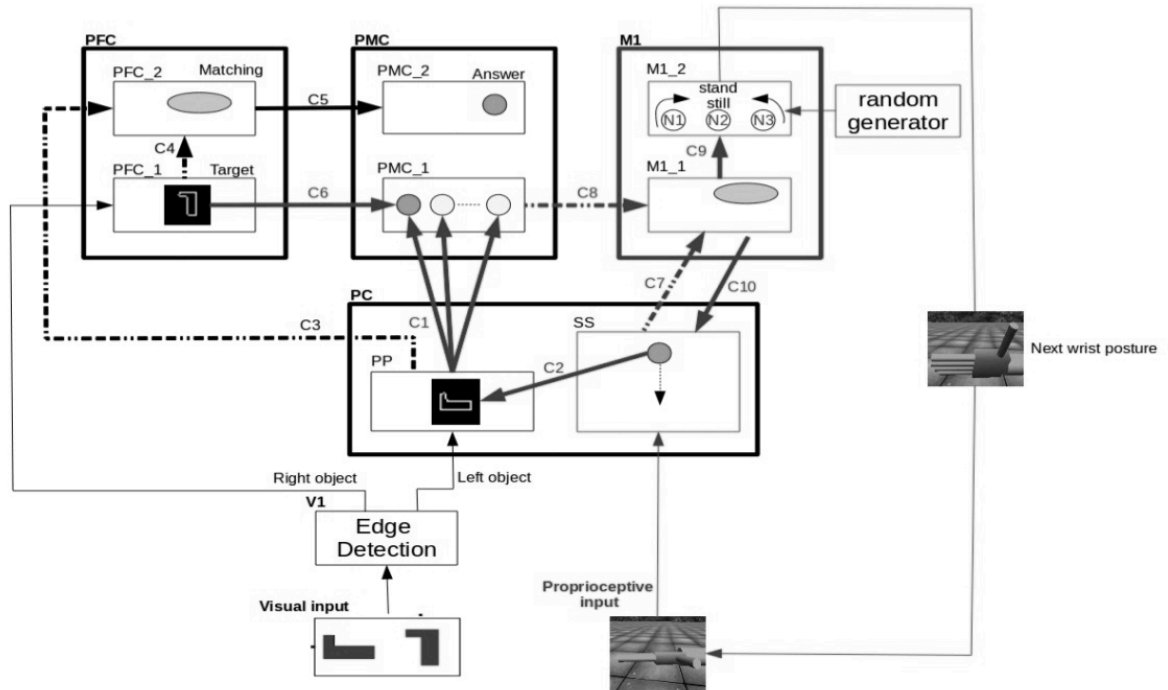


Figure 2. The model of mental image rotation. Each box represents the model's components. The arrows represent information flows from one component to another. The arrows accompanied by the letter "C" are the connections learned by SOM learning rule (dash-dot arrows) or by Hebbian learning rule (solid thick arrows).

The visual input for the model is the image of a simulated camera of one of the eyes of a simulated iCub robot. The edge information for the object on the left is passed to the PP, while the one for the “target object” on the right is for the PFC_1. The target object is used as a reference for rotational purposes. The robot has to mentally rotate the object encoded by PP to check if it is the same or it is different with respect to the target object stored within PFC_1. PFC_1 supplies also a bias signal to PMC_1 to lead the full activation (with a level of neural activation of 1.0) of one affordance among the elicited ones so transforming it into the representation of a desired wrist posture. This cluster represents the desired posture that the robot has to (mentally) reach to make a mental rotation useful to overlap the image within PP with the target image within PFC_1. PFC_2 is the core for the matching process. It is formed by a Kohonen self-organizing map (SOM) (Kohonen, 1997) which takes inputs from the PP and PFC_1. At the end of the matching process, PFC_2 neurons trigger PMC_2 activation whose neurons in turn encode the answering behaviour.

M1 consists of two areas M1_1 and M1_2. M1_1 is a SOM map (64 x 64 neurons) responsible for encoding a combination of current posture (from SS) and desired orientation (from PMC_1). The neural activation of M1_1 is used as input from SS during the mental rotation process (see below) as well as to trigger a wrist rotation through M1_2. This is a neural array formed by three (clusters of) neurons (N1, N2, N3). The activation of N1 causes a 30° clockwise rotation of the wrist; the activation of N3 causes a 30° counter-clockwise rotation of the wrist; the activation of N2 does not lead any rotation of the wrist.

2.2.2 The mental rotation process simulated by the model

This section briefly summarizes how the model reproduces the mental rotation processes. The follows points refer to the model functioning after the learning phase.

Affordances pre-activation (C1). The left object image encoded by PP neurons pre-activates all the possible object affordances within PMC_1 (wrist postures for a given object) at the same time. Since one object could assume 7 different orientations, we have 7 different clusters of neurons pre-activated within PMC_1. The affordances pre-activation mimics the preparatory processes for actions present when people see an object.

Affordances selection (C6). PFC_1 supplies a bias signal to PMC_1 to lead the full activation (with a level of neural activation of 1.0) of one affordance among the elicited ones so transforming it into the representation of a desired wrist posture. This cluster represents the desired posture that the robot has to (mentally) reach to make a

mental rotation useful to overlap the image within PP with the target image within PFC_1.

Mental rotation by the forward model and by the proprioceptive signal (C7, C8 and after this C10, C2). The desired wrist posture (encoded by PMC_1) and the current wrist posture (encoded by the SS) are combined within M1_1 (C7, C8). The SOM cluster within M1_1 works as a forward model with respect to SS evoking a Gaussian cluster within SS corresponding to the next wrist posture (C10). The SS map is also activated by current proprioceptive signal due to the wrist movement. The overall cluster within SS activates the new rotated image within PP (C2) causing a mental rotation.

In line with empirical evidence (Chu & Kita, 2008; Chu & Kita, 2011) the current proprioceptive signal affects the mental rotation processes as the overall activation of SS depends by both the signal from the forward model (C10) as well as by the current wrist proprioceptive signal (Figure 2). We assume that attention mechanisms might drive the system to be more focused on the mental rotation task rather than to the effect of its wrist movements. The effect of this attention focus is simulated by considering (within SS) the signal from the current proprioception weaker than the signal from the forward model.

2.2.3 Learning process

Connections between maps are trained using Hebbian learning and SOM competitive learning which are widely accepted as a biologically plausible learning mechanism mainly involving cortical areas (Doya, 2000). The specific Hebbian learning method used in this model is the Oja rule (Oja, 1982), a Hebbian like equation that solves the problem of the basic Hebb rule causing a weights growing without bound. The equations used to implement the Hebbian learning process are as follows:

$$\Delta w_{ij} = \eta a_i (a_j - w_{ij}); \quad w(t)_{ij} = w(t-1)_{ij} + \Delta w_{ij} \quad (1)$$

where Δw_{ij} denotes the weight's change from neuron i to neuron j , a_i and a_j denote activation potential of neuron i and j respectively, η denotes the learning rate which is set to 0.15, and $w(t)_{ij}$ is a weight value at a particular time step. The SOM learning rule has been implemented using the follow equation:

$$w(t)_i = w(t-1)_i + \Theta(t-1)_i \eta(t-1)_i (v(t-1)_i - w(t-1)_i) \quad (2)$$

where $w(t)_i$ denotes current weight value of neuron i at time t , $w(t-1)_i$ denotes an old weight value of the neuron i , Θ denotes the amount of influence on distance between neuron i and the best matching neuron in a map, η denotes the learning rate which is set to 0.15. Note that, Θ and η decrease over time. The Table 1 shows the parameters used for learning of the various connections.

Table 1. The parameters used in the network.

Connection	Type	number of patterns	Training cycles	Type of output
C1	Hebb	14	84	Cluster of activity
C2	Hebb	14	84	Image
C3 & C4	Kohonen	98	10,000	Cluster of activity
C5	Hebb	196	1,176	Cluster of activity
C6	Hebb	14	84	Cluster of activity
C7 & C8	Kohonen	196	10,000	Cluster of activity
C9	Hebb	98	1,960	Cluster of activity
C10	Hebb	98	1,960	Cluster of activity

Now we describe the training phases leading the model to perform the mental rotation task. The learning of the sensory-motor mapping (C7, C8, C9) and of the forward model (C10, C2) is accomplished at the beginning of the robot life. The aim of the sensory-motor learning phase is to obtain the values of the connection weights between SS-M1_1, between PMC_1-M1_1 and between M1_1-M1_2, useful to get a wrist rotation (encoded by M1_2) driving the current wrist posture (encoded by SS) towards the desired wrist posture (encoded by PMC_1).

The learning phase pivots on the follows “motor babbling procedure”: (a) the robot assumes a random wrist posture within $[-90^\circ, 90^\circ]$, which is encoded by a Gaussian cluster within SS; (b) the random generator randomly decides the direction of rotation (DR) and the number of rotations (NR). For example, if $DR = 1$ and $NR = 3$ the robot has to clockwise rotate its wrist of 90° ($3 \times 30^\circ$). $DR = 1$ causes the activation of the neuron N1 of M1_2. $NR = 3$ implies that N1 is activated for three sequential steps. We assume that “one time step” is the time the robot needs to rotate its wrist of 30° ; (c) the value of the wrist rotation is used to compute the total rotation (in this case $3 \times 30^\circ = 90^\circ$) and, based on the current posture, it is used to activate the PMC_1 map as a possible desired wrist posture; (d) PP neurons encode the current object orientation; (e) at the end of each step the Kohonen rule (2) is used to update the connection values (C7, C8) in order to obtain different cluster within the M1_1 representing all the combinations of the desired final wrist posture (PMC_1) and the current wrist posture (SS); (e) aside the SOM M1_1, at the end of each step we also train the forward model (C10, C2). Each SOM cluster (M1_1) is associated by the Hebbian rule (1) with the following wrist posture cluster (SS) which is in turn associated by (1) with the corresponding object orientation (PP) (this corresponds to perform a rotation with an object in the hand and associating the felt proprioception with the seen object image); (f) at the end each step the clusters activated within the SOM M1_1 are associated to M1_2 activated neuron (C9) using (1). The use of the SOM M1_1 is necessary to learn all the possible combinations between current posture (SS), desired posture (PMC_1), and control signal (M1_2). Overall there are 7 possible desired postured encoded in PMC_1 and 7×14 possible combinations to be encoded in M1_1.

Learning the affordances pre-activation (C1). The training pattern is constituted by 2 series of rotating images which are 30° degrees different per step. Each image is loaded in the PP area as activity level of a set of neurons in the map. The aim of the training process is to create a mapping between the input image (PP) and all the possible wrist posture of the robot encoded by cluster of activities (Gaussian tuning curve) within PMC_1. Importantly, the signal from PP pre-activates the clusters within PMC_1 with a value of 0.2 (the pre-activation is easily obtained by opportunely setting the max value of the PP-PMC_1 connection weights). This means that the object pre-activates several affordances (and not only one). The signal from PFC_1 allows the full activation, and hence the selection, of one cluster (one affordance) according the organism's goal (in our case the goal is given by the target image within PFC_1) (cf. Sec. 2.2). The training process is implemented using the Hebbian learning rule (1).

Learning the affordances selection (C6). The training pattern is formed by 2 series of rotating images which are 15° different per step. Each image is loaded into a PFC_1 map as activity level of a set of neurons in the map. An important difference with respect to the pre-activation of affordances training phase discussed above, is that here the aim of the training process is to create a mapping (by the Hebbian rule (1)) between specific target image (PFC_1) and specific wrist posture of the robot encoded by clusters of activities (Gaussian tuning curve) within PMC_1. In this way the signal from PP pre-activates within PMC_1 all the possible affordances (i.e. the 7 possible desirable wrist postures) related to the seen object, whereas the signal from PFC_1 supplies the crucial bias signal to select the desired wrist posture related to the target object.

Learning the matching and the answering processes (C3, C4, C5). The connections from PP and PFC_1 to SOM PFC_2 (C3, C4) are responsible for the matching process. When the network generates a mental image in the PP, having the same degrees rotation of the target image encoded by PFC_1, then the process of learning is triggered. The connections link two maps, one is PFC_1 (target image), which is set at the beginning of the simulation, and another is PP (the mental image). A training set for PFC_2 is a combination of all the possible neural representations for the stimuli of each input. A neural activity in PFC_2 forms a salient cluster with

respect to the two specific inputs. As there are 14 possible images in each map, four 196 clusters will be formed. To train PFC_2 has been used the SOM learning rule (2). The PFC_2 SOM map is trained in advanced. In this way, a response of PMC_2 can be fixed for each input couple from PP and PFC_1.

The answer triggering process uses the connection C5 from PFC_2 to PMC_2. When two images are “similar” the robot chooses the “YES” answer, otherwise it chooses the “NO” answer. The term “similar” means “it is approximately the same”. The mental rotation ends when the position of cluster of activity in M1_2 is close to the stand still position (N2). The most salient cluster in PFC_2 is used to produce the answer. Given the 196 possible combinations of inputs in the matching process, half of them are responsible for a “SAME” answer, while the remaining half for the “MIRROR” answer. Therefore, 98 regions in PFC_2 with respect to the same image from the PP and PFC_1 cause one cluster in PMC_2. While 98 other regions within PFC_2 represent different images of the two input maps. In this process, PMC_2 is responsible for the answer triggering, the motor response to press two answer buttons or to produce some utterance such as “YES” or “NO”. In the current version of the model this motor command is still not used to supply a control signal for the iCub but is directly interpreted as the response of the system.

After learning, an action potential of each neuron in the PMC_2 map is calculated by using a dynamic competition method (Erlhagen & Schöner, 2002). As the connections within a neural map are based on an all-to-all pattern, each neuron in the map sends/receives signals to/from every neuron. The dynamic competition process causes dynamic activities within the map, based on a distance between neurons following the rule of long-range inhibition and short-range excitation. Neighbouring neurons which are activated with high potential will receive excitatory signals and tend to form clusters of activity. In contrast, neurons which are far from the active neuron in the neural space will receive an inhibition signal and their action potential will be depressed.

The dynamic competition is also used as a method to calculate an agent’s response time (RT), e.g. to compare the model results with reaction time data in psychology experiments. Unlike a simple feed-forward process in layered neural networks, the dynamic competition process will be repeated until the action potential of at least one neuron in the neural map reaches a specific threshold. This process can be used to calculate the response time based on the action potential of an individual neuron that is most sensitive to a particular input. In detail, the number of repeating dynamic competition processes was recorded and used as simulated response time. One cycle of repeating the

process will be assumed to be equal to 1 millisecond (Caligiore et al., 2010).

3 Results

The two stimuli of the simulated mental rotation task can be varied in seven angular positions in the range $[-90^\circ; 90^\circ]$ with a step of 30° . Therefore the maximum angular disparity between the two stimuli is 180° and requires six rotational steps to mentally overlap the left stimulus to the target one. When the number of rotation cycles is equal to 10, it indicates that the model cannot correctly perform the mental image rotation of the left stimulus and will be forced to do the next step (matching process) by using the last image; the answer for the current mental rotation task (see sec. 2.1).

Figure 3(a) shows the mental rotation steps (PP) and the matching (PFC_2) and answering (PMC_2) processes for a successful trial. In this example the mental rotation process takes 5 steps to rotate an image of a stimulus in -60° to an image of stimulus in 90° , both stimuli are object-A. The mental rotation process ends when the rotated image reaches 90° orientation. After that, the matching process within PFC_2 is performed by using as input the neural activity of target image in PFC_1, and the rotated image in PP. The neural activation representing the matching process within PFC_2 is showed in the third column of the last row on Figure 3(a). The answering process of PMC_2, is indicated in the fourth column of the last row on figure 3a. The cluster of activity formed in the left side of the map will cause the answer “YES” to be chosen. The blank panels indicated that the rotational steps needed in this sample are less than 10.

In contrast, Figure 3(b) shows one case in which the model cannot rotate the left stimulus of 0° into the 60° position as the target stimulus. The model fails to rotate the image within 10 cycles, and has to do the matching process by using the last (un-rotated) image in PP. This scheme is similar to a guessing process in human subjects when the time to do mental rotation task is over. The model fails to rotate the image after 10 cycles: each cycle, the image in the PC is changed but cannot go to be the preferred image. This case might be caused by a mismatch cluster in SS that is also caused by noisy cluster’s position of M1_1. As indicates by the panel encoding, mental-0, and target in Figure 3(b), the direction of rotation is incorrect.

Possible failures in rotation and response of the model mainly come from the map M1_1 and connection C9 and C10. Because there are too many possible patterns to train these connections a cluster of active neurons in M1_1 might be overlapped in a position of clusters from other patterns and that overlapped might generate incorrect or noisy outputs. This feature of the model

simulates the error response often found in human subjects.

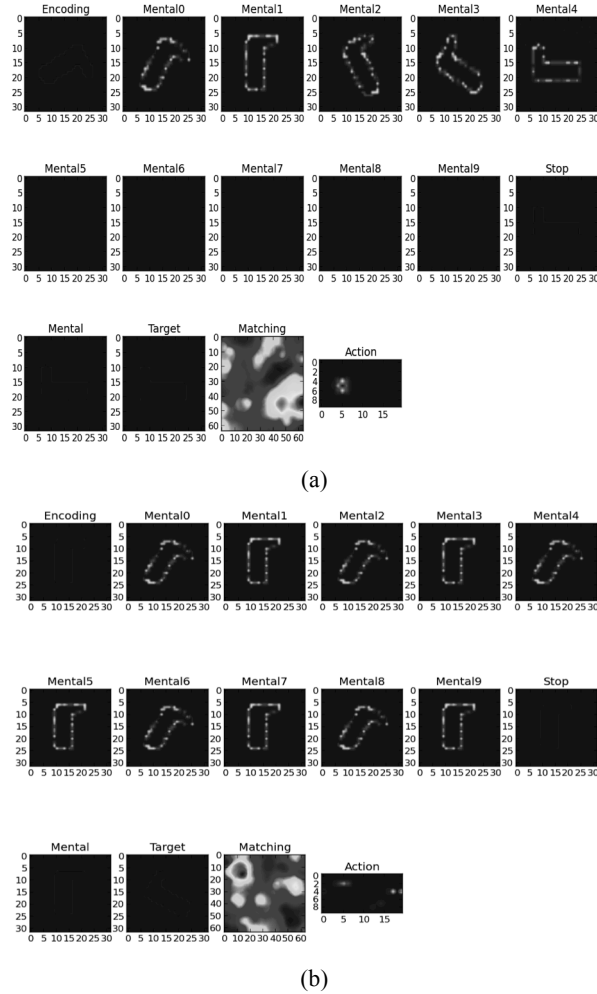


Figure 3. Mental image rotation steps. Where (a) indicates rotational steps in the case that the model is able to create a series of image changes to reach the target orientation; (b) the model is unable to rotate the seen object. The matching and answering processes are represented by the neural activation of the two bottom right side maps respectively.

After testing the model with all possible pairs of stimuli used in the training set, the model achieve 97.95% (192 out of 196) successful rate of rotation of the left stimulus to match with the target. The overall percentage of correct response (answer triggering) is 85.7% (168 out of 196).

As indicated by RTs profiles showed in Figure 4(a), when the angular disparity is high, the required cycle of rotation and RTs also increase. The angular disparity (x-axis) is calculated by using the difference in orientation between the two stimuli. The 0° disparity is calculated from the left stimulus orientation that is the same as the right one but can be from the different type of object. As indicates by the RTs profile, there is no significantly effect from the different types of object that are used in rotation and comparison process.

There are three types of errors that have been generated by uncontrolled situations within the model. The first is from the situation that the model cannot rotate the left stimulus to match with the right one within 10 rotational cycles. An error from this case will cause higher response time than normal cases, and also the cause of incorrect response. Secondly, as the connection from SS to PP underlies a forward model, a possible positioning error in SS directly effects a mental image in PP. Therefore, it raises the issue of successful rotation by accident. In detail, when active neurons in M1_1 cause an incorrect cluster in SS, by chance, it might be a cluster that causes an image of the target of a given task. And then the model stops rotational processes. In this case, the number of rotation will be less than usual which effects response times and also might causes an incorrect answer. Lastly, even the model can correctly and successfully rotate the left stimulus of a given task, but the answer might not be always correct. The first two errors are caused by neuron activity within the map M1_1 while the last error is from the map PFC_2.

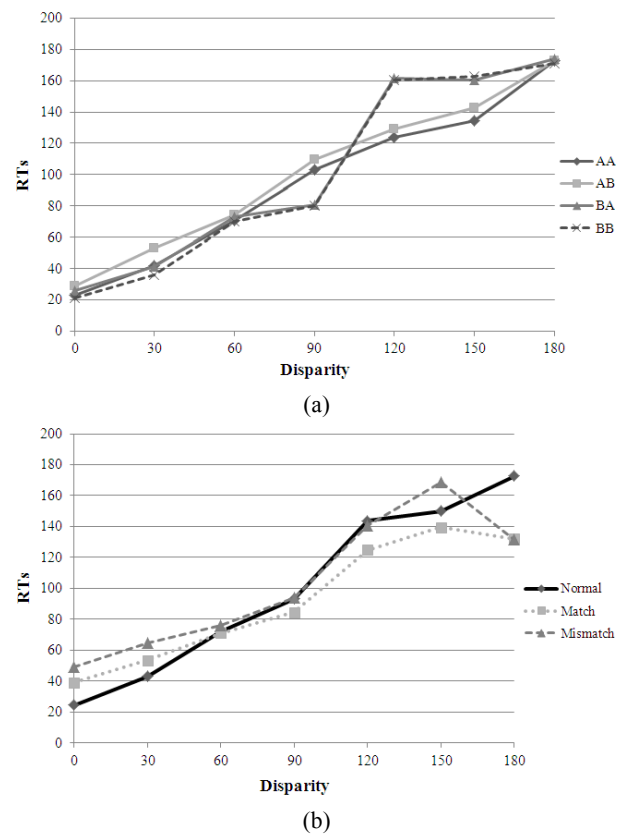


Figure 4. The comparison of response time profiles. Where panel (a) indicates the difference of response time profile between different pairs of stimuli (b) shows the difference between supplied a match/mismatch of proprioceptive signal to SS. Label AA in the panel (a) denotes that the left stimulus is object-A and compare with the target that is object-A, while AB, BA, and BB are used to denote each specific line in the graph by the same meaning as AA.

The proprioceptive signal has been simulated by using the current wrist's angle of the robot. A specific cluster in SS is involved according to the wrist's angle. This process acts as a cluster pre-activation for the map SS. When the position of the pre-active cluster and of the cluster caused by M1_1 are the same or overlapped, it should support the rotational processes and the response times will be reduced which means the model could do rotation faster. In contrast, if they are difference, the dynamic competition processes should take a longer time to activate the most salient neuron within a map. The match situation between the current wrist's angle and the felt proprioception (SS clusters) is simply simulated by activating the overlapping clusters as described above. On the other end, the mismatch situation is caused by the chance activation of a cluster in SS (using random number generator to choose a specific cluster). The response time profiles in Figure 4(b) indicate the difference between proprioceptive signal match and mismatched with the felt proprioception for object's orientation.

4 Conclusion

The neurorobotics model proposed in this paper accounts for the mental rotation processes based on neural mechanisms involving visual imagery, affordance encoding and forward models processing. In this respect, the proposed approach is in agreement with the most recent theoretical and empirical research on mental rotation (Lamm et al., 2007).

Importantly, in addition to replicating the typical mental rotation data, the model is able to account for other data which link overt movements and mental rotations (Wohlschläger & Wohlschläger, 1998; Wohlschläger, 2001) (cf. Introduction). This recent empirical evidence claims that the performance of mental rotation tasks can be improved by the assistance of hand movements, or gestures called “co-thought gestures” (Chu & Kita, 2008; Chu & Kita, 2011). Spontaneous gestures during the performance of mental rotation provide a rich sensorimotor experience to the solving strategy in human subjects. Gestures improve the internal representation of a spatial transformation of objects. Following this hypothesis, the proposed model includes proprioceptive units that act as an internal representation of wrist movements, which directly affect the mental rotation process within the parietal-premotor circuit. Interestingly, the model suggests an operational hypothesis on how the covert mental rotation neural mechanisms pivoting of forward model circuit and overt movements might be combined to affect mental rotation.

The model is validated within the simulated humanoid robot iCub engaged in solving a mental rotation tasks. The provides a demonstration that the integration of

mental rotation capabilities with the affordance and embodiment processes (developed in the motor babbling training phase) leads to the successful performance of mental rotation task.

This sets the basis for ongoing work on the extension of this model for investigating the role of co-thought gestures (Chu & Kita, 2008; Chu & Kita, 2011) to support mental rotation tasks, as well as other cognitive capabilities as for communicative gesture use and language learning.

Future extension of this model will also look at the use of a variety of objects for mental rotation. The current model can only process and rotate the objects of the training set. To permit the rotation of unseen objects, the object orientation mechanisms will be separated by the object identity function, e.g. using an inferotemporal cortex (IT) map whose neurons encode objects independently of their orientation.

Overall the proposed neurorobotic model of mental rotation provides a useful computational framework to study the integration between mental rotation capabilities and embodied cognition, to demonstrate the role of motor processes and affordance in mental simulation task.

Funding

This research received funds from EU funded projects Robot-Era (FP7-IP-288899), Poeticon++ (FP7-STREP-288382), a grant from the Thai government, and “IM-CLeVeR–Intrinsically Motivated Cumulative Learning Versatile Robots”, contract no. FP7-IP -231722.

References

- Cangelosi, A., & Schlesinger, M. (in press) *Developmental robotics*, MIT press.
- Caligiore, D., Borghi, A.M., Parisi, D., & Baldassarre, G. (2010). TRoPICALS: A computational embodied neuroscience model of compatibility effects. *Psych Rev*, 117, 1188-1228.
- Caligiore, D., Borghi, A.M., Parisi, D., Ellis, R., Cangelosi, A., & Baldassarre, G. (2012). How affordances associated with a distractor object affect compatibility effects: A Study with the computational model TRoPICALS. *Psych Res*, DOI: 10.1007/s00426-012-0424-1
- Canny, F. J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8, 6, 679-698.
- Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: insights into the microdevelopment of the motor strategy. *J. Exp. Psychol. Gen*, 137, 706-723.

- Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *J. Exp. Psychol. Gen.*, 140, 102-116.
- Corballis, M.C., & McLaren R. (1982). Interaction between perceived and imagined rotation. *J Exp Psychol Hum Percept Perform*, 8, 215-224.
- Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology*, 10, 732-739.
- Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review*, 109, 3, 545-572.
- Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks*, 11, 7-8, 1277-1303.
- Fuster, J.M. (2001). The prefrontal cortex--an update: time is of the essence. *Neuron*, 30, 2, 319-333.
- Georgopoulos, A.P., Lurito, J.T., Petrides, M., Schwartz, A.B., & Massey, J.T. (1989). Mental rotation of the neuronal population vector. *Science*, 243, 234-236.
- Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston, Houghton-Mifflin.
- Grafton, S.T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *NeuroImage*, 6, 231-236.
- Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp*, 121, 1-19.
- Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioural and Brain Sciences*, 27, 3, 377-396.
- Kohonen, T. (1997). *Self-organizing maps*, 2nd ed. Berlin, Germany: Springer-Verlag.
- Kosslyn, S.M. (1994). *Image and the Brain. The Resolution of the Imagery Debate*. MIT Press/A Bradford Book, Cambridge
- Lamm, C., Fischmeister P.H.S., & Bauer, H. (2005). Individual differences in brain activity during visuo-spatial processing assessed by slow cortical potentials and LORETA. *Cogn Brain Res*, 25, 900-912.
- Lamm, C., Windischberger, C., Moser, E., & Bauer, H. (2007). The functional role of the dorso-lateral premotor cortex during mental rotation An event-related fMRI study separating cognitive processing steps using a novel task paradigm. *NeuroImage*, 36, 1374-1386.
- Miall, R. C. (2003). Connecting mirror neurons and forward models. *Neuroreport*, 14, 17, 2135-2137.
- Oja, E. (1982). A simplified neuron model as a principal component analyser. *Journal of Mathematical Biology*, 15, 267-173.
- Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, 26, 381-410.
- Richter, W., Somorjai, R., Summers, R., Jarmasz, M., Menon, R.S., & Gati, J.S. (2000). Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *J Cog Neurosci*, 12, 310-320.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review Neuroscience*, 27, 169-192.
- Sandini, G., Metta, G., & Vernon, D. (2007). The iCub cognitive humanoid robot: An open-system research platform for enactive cognition. in *50 Years of AI, M. Lungarella, Ed. et al.* Berlin, Germany: Springer-Verlag, 359-370.
- Shepard, R., & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science*, 171(972):701-3.
- Tikhonoff, V., Cangelosi, A., Fitzpatrick, P. Metta, G., Natale, L., & Nori, F. (2008). An open-source simulator for cognitive robotics research: The prototype of the iCub humanoid robot simulator. in *Proc. IEEE Workshop*.
- Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Vis Cogn*, 8, 769-800.
- Wexler, M., Kosslyn, S.M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, 68, 1, 77-94.
- Wohlschläger, A. & Wohlschläger, A. (1998). Mental and manual rotation. *J Exp Psychol Hum Percept Perform*, 242, 397-412.
- Wohlschläger, A. (2001). Mental object rotation and the planning of hand movements. *Percept Psychophys*, 63, 709-718.
- Wolpert, D.M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Netw*, 11, 1317-1329.